

Name: Santos, Tristan Neal U.  
Course and Section: CPE32S9  
Date of Submission: 02-07-2024  
Instructor: Engr. Roman Richard


#### Part 1, Step 1:

```
from google.colab import drive
drive.mount('/content/drive')
```

Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force\_remount=True).

#Code Cell 1

```
import pandas as pd
brainFile = '/content/drive/MyDrive/Colab Notebooks/Emtech2/Text/brainsize.txt'
brainFrame = pd.read_csv(brainFile, '\t')
```

 <ipython-input-20-b496ce176638>:4: FutureWarning: In a future version of pandas all arguments of read\_csv except for the argument 'filepath\_or\_buffer' will be keyword-only.  
brainFrame = pd.read\_csv(brainFile, '\t')

#### Step 2

#Code Cell 2

```
brainFrame.head()
```

	Gender	FSIQ	VIQ	PIQ	Weight	Height	MRI_Count
0	Female	133	132	124	118.0	64.5	816932
1	Male	140	150	124	NaN	72.5	1001121
2	Male	139	123	150	143.0	73.3	1038437
3	Male	133	129	128	172.0	68.8	965353
4	Female	137	132	134	147.0	65.0	951545

#### Part 2, Step 1:

#Code Cell 3

```
brainFrame.describe()
```

	FSIQ	VIQ	PIQ	Weight	Height	MRI_Count
<b>count</b>	40.000000	40.000000	40.000000	38.000000	39.000000	4.000000e+01
<b>mean</b>	113.450000	112.350000	111.02500	151.052632	68.525641	9.087550e+05
<b>std</b>	24.082071	23.616107	22.47105	23.478509	3.994649	7.228205e+04
<b>min</b>	77.000000	71.000000	72.00000	106.000000	62.000000	7.906190e+05
<b>25%</b>	89.750000	90.000000	88.25000	135.250000	66.000000	8.559185e+05
<b>50%</b>	116.500000	113.000000	115.00000	146.500000	68.000000	9.053990e+05
<b>75%</b>	135.500000	129.750000	128.00000	172.000000	70.500000	9.500780e+05
<b>max</b>	144.000000	150.000000	150.00000	192.000000	77.000000	1.079549e+06

Step 2, A:

```
#Code Cell 4
import numpy as np
import matplotlib.pyplot as plt
```

Step 2, B:

```
#Code Cell 5
menDf = brainFrame[(brainFrame.Gender == 'Male')]
womenDf = brainFrame[(brainFrame.Gender == 'Female')]
womenDf.head()
```

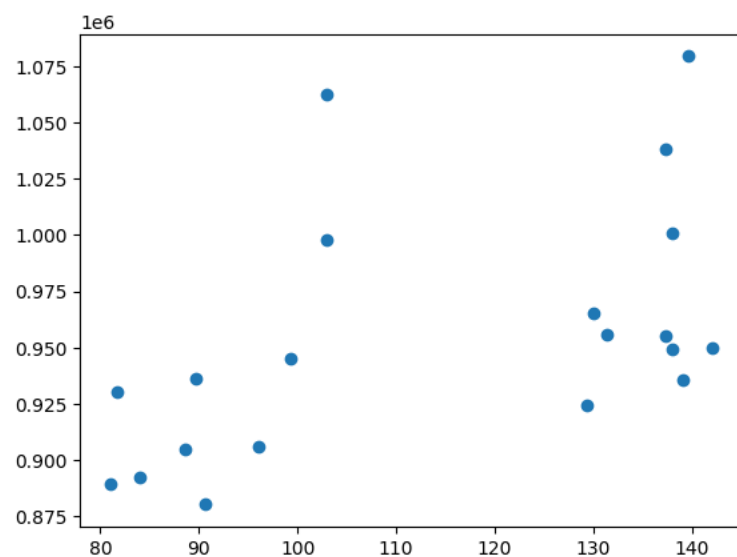
	Gender	FSIQ	VIQ	PIQ	Weight	Height	MRI_Count
<b>0</b>	Female	133	132	124	118.0	64.5	816932
<b>4</b>	Female	137	132	134	147.0	65.0	951545
<b>5</b>	Female	99	90	110	146.0	69.0	928799
<b>6</b>	Female	138	136	131	138.0	64.5	991305
<b>7</b>	Female	92	90	98	175.0	66.0	854258

```
#Code Cell 5
menDf = brainFrame[(brainFrame.Gender == 'Male')]
womenDf = brainFrame[(brainFrame.Gender == 'Female')]
menDf.head()
```

	Gender	FSIQ	VIQ	PIQ	Weight	Height	MRI_Count
1	Male	140	150	124	NaN	72.5	1001121
2	Male	139	123	150	143.0	73.3	1038437
3	Male	133	129	128	172.0	68.8	965353
8	Male	89	93	84	134.0	66.3	904858
9	Male	133	114	147	172.0	68.8	955466

Step 2, C:

```
# Code cell 6
menMeanSmarts = menDf[["PIQ", "FSIQ", "VIQ"]].mean(axis=1)
plt.scatter(menMeanSmarts, menDf["MRI_Count"])
plt.show()
%matplotlib inline
```



```
# Code cell 7
# Graph the women-only filtered dataframe
#womenMeanSmarts = ?
#plt.scatter(?, ?)
plt.show()
%matplotlib inline
```

Part 3, Step 1:

```
# Code cell 8
brainFrame.corr(method='pearson')
```

```
<ipython-input-33-cab48f3abe05>:2: FutureWarning: The default value of numeric_only in DataFrame.corr is deprecated. In a future version, it will default to False. Select only valid columns
brainFrame.corr(method='pearson')
```

	FSIQ	VIQ	PIQ	Weight	Height	MRI_Count
<b>FSIQ</b>	1.000000	0.946639	0.934125	-0.051483	-0.086002	0.357641
<b>VIQ</b>	0.946639	1.000000	0.778135	-0.076088	-0.071068	0.337478
<b>PIQ</b>	0.934125	0.778135	1.000000	0.002512	-0.076723	0.386817
<b>Weight</b>	-0.051483	-0.076088	0.002512	1.000000	0.699614	0.513378
<b>Height</b>	-0.086002	-0.071068	-0.076723	0.699614	1.000000	0.601712
<b>MRI_Count</b>	0.357641	0.337478	0.386817	0.513378	0.601712	1.000000

```
# Code cell 9(Female)
```

```
womenDf.corr(method='pearson')
```

```
<ipython-input-34-a6271751808a>:2: FutureWarning: The default value of numeric_only in DataFrame.corr is deprecated. In a future version, it will default to False. Select only valid columns
womenDf.corr(method='pearson')
```

	FSIQ	VIQ	PIQ	Weight	Height	MRI_Count
<b>FSIQ</b>	1.000000	0.955717	0.939382	0.038192	-0.059011	0.325697
<b>VIQ</b>	0.955717	1.000000	0.802652	-0.021889	-0.146453	0.254933
<b>PIQ</b>	0.939382	0.802652	1.000000	0.113901	-0.001242	0.396157
<b>Weight</b>	0.038192	-0.021889	0.113901	1.000000	0.552357	0.446271
<b>Height</b>	-0.059011	-0.146453	-0.001242	0.552357	1.000000	0.174541
<b>MRI_Count</b>	0.325697	0.254933	0.396157	0.446271	0.174541	1.000000

```
# Code cell 10
```

```
# Use corr() for the male-only dataframe with the pearson method
```

```
#?.corr(?)
```

```
# Code cell 9(Male)
```

```
menDf.corr(method='pearson')
```

```
<ipython-input-36-9c6b27fe7b04>:2: FutureWarning: The default value of numeric_only in DataFrame.corr is deprecated. In a future version, it will default to False. Select only valid columns
menDf.corr(method='pearson')
```

	FSIQ	VIQ	PIQ	Weight	Height	MRI_Count
<b>FSIQ</b>	1.000000	0.944400	0.930694	-0.278140	-0.356110	0.498369
<b>VIQ</b>	0.944400	1.000000	0.766021	-0.350453	-0.355588	0.413105
<b>PIQ</b>	0.930694	0.766021	1.000000	-0.156863	-0.287676	0.568237
<b>Weight</b>	-0.278140	-0.350453	-0.156863	1.000000	0.406542	-0.076875
<b>Height</b>	-0.356110	-0.355588	-0.287676	0.406542	1.000000	0.301543
<b>MRI_Count</b>	0.498369	0.413105	0.568237	-0.076875	0.301543	1.000000

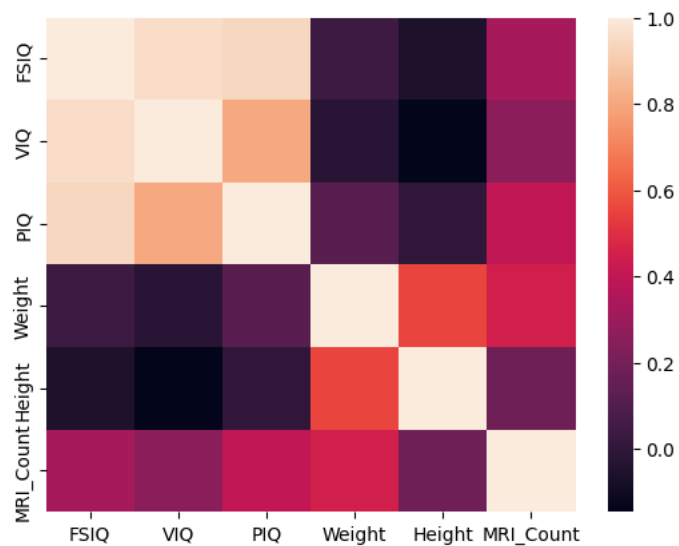
#### Part 4, Step 1:

```
# Code cell 11
!pip install seaborn
```

```
Requirement already satisfied: seaborn in /usr/local/lib/python3.10/dist-packages (0.13.1)
Requirement already satisfied: numpy!=1.24.0,>=1.20 in /usr/local/lib/python3.10/dist-packages (from seaborn) (1.23.5)
Requirement already satisfied: pandas>=1.2 in /usr/local/lib/python3.10/dist-packages (from seaborn) (1.5.3)
Requirement already satisfied: matplotlib!=3.6.1,>=3.4 in /usr/local/lib/python3.10/dist-packages (from seaborn) (3.7.1)
Requirement already satisfied: contourpy>=1.0.1 in /usr/local/lib/python3.10/dist-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (1.2.0)
Requirement already satisfied: cycler>=0.10 in /usr/local/lib/python3.10/dist-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in /usr/local/lib/python3.10/dist-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (4.47.2)
Requirement already satisfied: kiwisolver>=1.0.1 in /usr/local/lib/python3.10/dist-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (1.4.5)
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.10/dist-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (23.2)
Requirement already satisfied: pillow>=6.2.0 in /usr/local/lib/python3.10/dist-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (9.4.0)
Requirement already satisfied: pyparsing>=2.3.1 in /usr/local/lib/python3.10/dist-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (3.1.1)
Requirement already satisfied: python-dateutil>=2.7 in /usr/local/lib/python3.10/dist-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.10/dist-packages (from pandas>=1.2->seaborn) (2023.4)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.10/dist-packages (from python-dateutil>=2.7->matplotlib!=3.6.1,>=3.4->seaborn) (1.16.0)
```

```
# Code cell 12
import seaborn as sns
wcorr = womenDf.corr()
sns.heatmap(wcorr)
plt.savefig('attribute_correlations.png', tight_layout=True)
```

```
<ipython-input-38-2465c40f5efb>:3: FutureWarning: The default value of numeric_only in DataFrame.corr is deprecated. In a future version, it will default to False. Select only valid c
wcorr = womenDf.corr()
<Axes: >
```



```
# Code cell 14
mcorr = menDf.corr()
sns.heatmap(mcorr)
plt.savefig('attribute_correlations.png', tight_layout=True)
```

```
<ipython-input-39-ff3e250059fc>:2: FutureWarning: The default value of numeric_only in DataFrame.corr is deprecated. In a future version, it will default to False. Select only valid columns.
mcorr = menDf.corr()
```

## Many variable pairs present correlation close to zero. What does that mean?

-Due to having weak and differential relations towards variable pairs to one another that it feels like theres no difference between them.

## Why seperate the Genders?

-To find some differentiation between them while also showcasing what output results we got from them.

## What variables have stronger correlation with brain size (MRI\_Count)? Is that expected? Explain.

-



## Supplementary

```
from google.colab import drive
drive.mount('/content/drive')
```

```
#Code Cell 1
import pandas as pd
dataFile = '/content/drive/MyDrive/Colab Notebooks/Emtech2/Text/buddymove (1).txt'
dataFrame = pd.read_csv(dataFile, sep='\t')
```

```
#Code Cell 2
dataFrame.head()
```

	User Id	Unnamed: 1	Sports	Unnamed: 3	Religious	Nature	Theatre	Shopping	Picnic	Unnamed: 9	
User 1	NaN	2	NaN	77	NaN	79	69	68	NaN	95	NaN
User 2	NaN	2	NaN	62	NaN	76	76	69	NaN	68	NaN
User 3	NaN	2	NaN	50	NaN	97	87	50	NaN	75	NaN